1 | 3–4 | 98

C tics and Cell Genetics

S. Karger
Medical and Scientific
Publishers
Basel • Freiburg
Paris • London
New York • New Delhi
Bangkok • Singapore
Tokyo • Sydney

KARGER

Applicants: David J. Pinsky
U.S. Serial No.: 09/374,586
Filed: August 13, 1999
Group Art Unit: 1633

Exhibit 33

Cytogenetics and Cell Genetics

Cytogenet Cell Genet
ISSN 0301–0171
CGCGBR
81(3–4) 169–302 (1998)

Chromosome Workshop

Original Articles

Gene Mapping, Cloning, and Sequencing

(Continued on reverse side)

KARGER

Cytogenet Cell Genet 81:287–289 (1998)

Cytogenetics and Cell Genetics

Gene structure and chromosome location of mouse *Cd39* coding for an ecto-apyrase

M.A. Schoenborn,[a] N.A. Jenkins,[b] N.G. Copeland,[b] D.J. Gilbert,[b] R.B. Gayle III,[a] and C.R. Maliszewski[a]

[a] Immunex Corporation, Seattle, WA;
[b] Mammalian Genetics Laboratory, NCI-Frederick, Cancer Research and Development Center, Frederick, MD (USA)

Rationale and significance

CD39 is a membrane-bound 95-kDa glycoprotein that exhibits potent ATP diphosphohydrolase (ATPDase) activity (Kaczmarek et al., 1996; Marcus et al., 1997). Its expression on vascular endothelial cells and the ability of recombinant human CD39 to block ADP-dependent platelet aggregation illustrate the potential importance of CD39 in thromboregulatory processes (Marcus et al., 1997). The cloning of human and mouse CD39 cDNAs and the chromosome location of human CD39 (10q23.1 →q24.1) have been described (Maliszewski et al., 1994). Considerable amino acid sequence homology exists between CD39 and NTPases from a phylogenetically diverse array of organisms (Handa and Guidotti, 1996). Homology is strongest within several discrete "apyrase conserved regions" (ACRs), which are probably required for enzymatic activity. To further our understanding of the relationship of members of the NTPase family we sought to elucidate the structure of *Cd39*.

Materials and methods

The chromosome location of *Cd39* was determined by interspecific backcross analysis using progeny derived from matings of (C57BL/6J × *Mus spretus*)F_1 females and C57BL/6J male mice as described (Copeland and Jenkins, 1991). A total of 205 N_2 mice were used to map the *Cd39* locus. This interspecific backcross mapping panel has been typed for over 2500 loci that are well distributed among all the autosomes, as well as the X-chromosome (Copeland and Jenkins, 1991). C57BL/6J and *M. spretus* DNAs were digested with several endonucleases and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) using a ^{32}P-labeled probe nick-translated from a 2.3-kb *Eco*RI fragment of mouse *Cd39* cDNA. Major fragments of 21.0, 5.6, 4.5 and 4.2 kb were detected in *Kpn*I digested C57BL/6 DNA, and major fragments of 14.0 and 9.8 kb were detected in *Kpn*I digested *M. spretus* DNA. The 14.0- and 9.8-kb *Kpn*I RFLPs from *M. spretus* were used to follow the segregation of the *Cd39* locus in backcross mice. A description of the probes and RFLPs for the loci linked to *Cd39* has been reported previously (Copeland et al., 1993). Recombination distances were calculated using Map Manager, version 2.6.5.

Hybridization of a mouse (strain 129/SV) genomic/λ-phage library with a ^{32}P-labeled mouse *Cd39* full-length cDNA probe resulted in the isolation of 12 clones. Restriction mapping, Southern analysis and DNA sequencing led to the identification of three overlapping contiguous genomic clones (gcD, gcF and gcH) and one nonoverlapping clone (gcX) that, collectively, contained the entire cDNA sequence. These four clones underwent additional DNA sequencing, PCR analysis, and restriction mapping.

Results and discussion

Mapping showed that *Cd39* is located in the central region of mouse chromosome 19 in linkage with *Fas*, *Tdt*, and *Col17a1* (Copeland et al., 1993). Although 64 mice were analyzed for every marker and are shown in the segregation analysis (Fig. 1), up to 113 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using the additional data. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci, and the most likely gene order are: centromere – *Fas* – 10/113 – *Cd39* – 0/112 – *Tdt* – 5/101 – *Col17a1*. The recombination frequencies are: – *Fas* – 8.9cM ± 2.7 – [*Cd39*,*Tdt*] – 5.0 cM ± 2.2 – *Col17a1*. No recombinants were detected between *Cd39* and *Tdt* in 112 animals typed in common, suggesting that the two loci are within 2.7 cM of each other (upper 95% confidence limit).

Supported in part by the National Cancer Institute, DHHS, under contract with ABL.

Received 11 March 1998; manuscript accepted 15 April 1998.

Request reprints from Michael A. Schoenborn, Immunex Corporation, 51 University St., Seattle, WA 98101 (USA); telephone: 206-587-0430; fax: 206-233-9733; e-mail: schoenborn@immunex.com

KARGER
Fax + 41 61 306 12 34
E-mail karger@karger.ch
www.karger.com

0301–0171/98/0814–0287$15.00/0

Accessible online at:
http://BioMedNet.com/karger

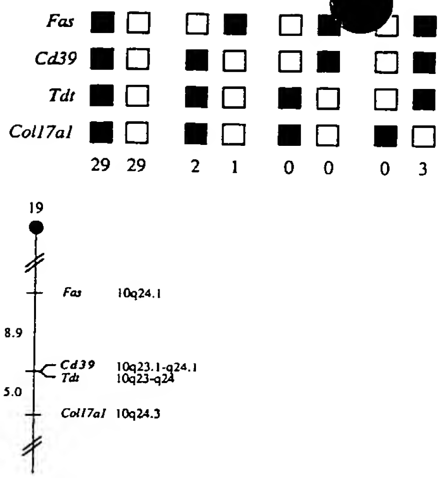

Fig. 1. *Cd39* was mapped in the central region of mouse chromosome 19 by interspecific backcross analysis. The segregation patterns of *Cd39* and flanking genes in 64 backcross animals that were typed for all loci are shown at the top of the figure. Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*)F_1 parent. The shaded boxes represent the presence of a C57BL/6J allele and white boxes represent the presence of an *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 19 linkage map showing the location of *Cd39* in relation to linked genes is shown at the bottom of the figure. Recombination distances between loci in centimorgans are shown to the left of the chromosome and the positions of loci in human chromosome, where known, are shown to the right. References for the human map positions of loci cited in this study can be obtained from the GDB (Genome Data Base at http://gdbwww.gdb.org/), a computerized database of human linkage information maintained by The William H. Welch Medical Library of Johns Hopkins University (Baltimore, MD).

We have compared our interspecific map of chromosome 19 with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (provided from Mouse Genome Database, a computerized database maintained at The Jackson Laboratory, Bar Harbor, ME). *Cd39* mapped in a region of the composite map that lacks mouse mutations with a phenotype that might be expected for an alteration in this locus (data not shown).

The central region of mouse chromosome 19 shares a region of homology with human chromosome 10q (summarized in Fig. 1). The placement of *Cd39* in this interval in mouse is consistent with the human localization of CD39 at 10q23.1→q24.1 (Maliszewski et al., 1994). Linkage analysis and allele loss studies have shown that this region of human chromosome 10 contains genes for prostate tumor suppressor, spinocerebellar ataxia, Cowden disease, development (PAX2, HOX11, and WNT8B), cytochrome P450IIC, and audiogenic partial epilepsy (Gray et al., 1997; Wang et al., 1997).

Analysis of the genomic clones revealed that *Cd39* spans at least 47.4 kb and consists of ten coding exons separated by nine introns (Fig. 2, Genbank Accession Nos. AF041812 through AF041818). Based on analysis of the content and order of coding exons, the predicted mRNA sequence from *Cd39* agrees with the cDNA sequence (GenBank Accession No. AF037366). Sizes were accurately determined for all but the first intron. Splice junctions were identified based upon comparison of genomic to cDNA sequence, and confirmed by identity with consensus sequences. Notably, introns II and VI each contain non-canonical GC variant splice donors. In both cases, the nucleotides at U1 RNA complementary positions match prototypic 5′ splice site bases (Jackson, 1991).

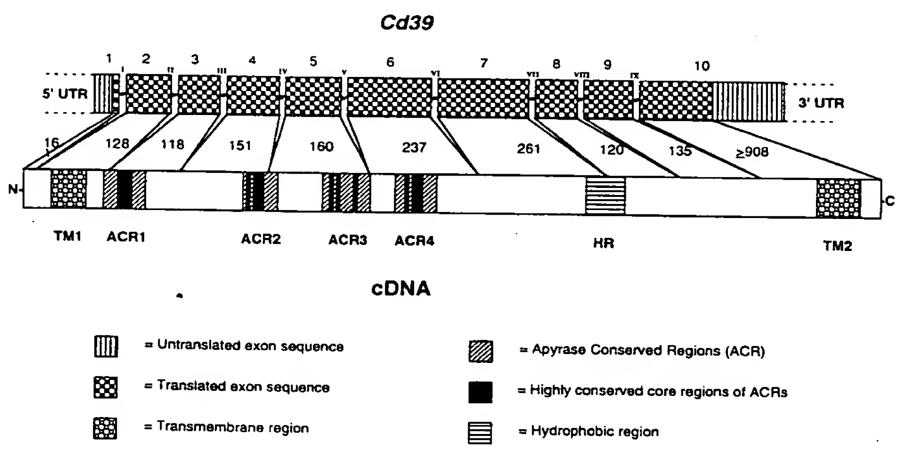

Fig. 2. Structural organization of *Cd39* in relation to functional cDNA domains. The genomic structure is shown with exons represented as numbered boxes. Within exons, translated and untranslated (UTR) segments are distinguished (See key). Sizes (in base pairs) are given for coding portions of exons and appear between genomic and cDNA. Intron designations appear as Roman numerals above respective condensed introns. Intron sizes may be found in Table 1. Correspondence between translated exons in genomic DNA (top) and coding cDNA (below) is indicated. ACRs, ACR core regions, transmembrane domains (TM), and the third hydrophobic region (HR) are depicted as patterned boxes (See key).

Table 1. Exon-intron junctions of *Cd39*

Exon	Splice donor	Intron	Splice acceptor	Exon	Intron Size	Interruption at or between codon(s) for amino acid(s)
1	A AAG Gg taagaccag	I	tttgactttagAT TCT	2	>19.6 kb	Asp^{6}
2	TT AAGgcaagtaaaa	II	ttgcccteagTAT GG	3	8.3 kb	Lys^{48}/Tyr^{49}
3	G AAA Gg taagacggg	III	cttccctca gGT CCT	4	2.6 kb	Gly^{88}
4	CTT AGg tagagtgat	IV	ttttgtgta gA ATG G	5	1687 bp	Arg^{138}
5	CT CAGg taaagacct	V	ttctttca gGAA CA	6	938 bp	Gln^{191}/Glu^{192}
6	TT CAGgcaagtgcaa	VI	ttccattca gGTT TC	7	1029 bp	Gln^{270}/Val^{271}
7	TT GGGg taagtttgt	VII	tgttccaca gGCG TT	8	6.0 kb	Gly^{357}/Ala^{358}
8	AA GAGg taagtgac	VIII	tttcttta gACA AA	9	3.0 kb	Glu^{397}/Thr^{398}
9	GC AAGg taacttggg	IX	ttctgttaaca gATC AA	10	2.0 kb	Lys^{442}/Ile^{443}

[a] Exon sequences appear in uppercase with codon nucleotides grouped. Intron sequences appear in lower case with splice consensus sequences (Jackson, 1991) underlined. Bold type sequences indicate identity with corresponding CD39-like-1 splice sites (Chadwick et al., 1997). Sizes were accurately determined for all but the first exon.

The organization of exon-intron structure in *Cd39* and the correspondence of exonic translated regions to cDNA domains are depicted in Fig. 2. Each of the transmembrane (TM) regions is fully contained within a discrete exon: 2 and 10, respectively. The third hydrophobic region is divided between exons 7 and 8. The four ACRs are encoded on four contiguous exons: 3 through 6, sequentially. None of the highly conserved core amino acid codons of the ACRs (Handa and Guidotti, 1996) is interrupted by an intron. Due to the apparent functional importance of the ACRs, these findings suggest that introns III, IV and V arose after an archetypal NTPase gene.

The cloning of the human CD39-like-1 gene (CD39L1) was based upon homology to CD39 cDNA. The structure of CD39L1 and the deduced amino acid sequence have been described (Chadwick and Frischauf, 1997). Comparison of our findings with those of Chadwick et al. indicate that striking structural similarities exist between these two genes. In particular, CD39L1 exons 1 through 7 are similar or identical in size to *Cd39* exons 2 through 8. The codon content, order of corresponding exons, and splice junction DNA sequences are also highly conserved. The corresponding intron sizes for the two genes are dissimilar. In contrast to demonstrated similarity in gene structure, the amino acid sequence of human CD39L1 is significantly more similar to chicken muscle ecto-ATPase (Kirley, 1997) and rat ecto-ATPase (Kegel et al., 1997) than it is to CD39. As such, it is likely that CD39L1 is a member of the ecto-ATPase subgroup, instead of the ATPDase subgroup of the NTPase family to which CD39 belongs (Kegel et al., 1997).

A search of the approximately 1.2 kb of 5′-flanking sequence from gcX (AF041812) using UW-GCG Transfac software revealed numerous transcription factor binding site motifs including TATA- and CCAAT-box promotor motifs (data not shown). A search of the non-redundant GenBank database showed that the 5′-flanking sequence has no significant matches with non-CD39 genes. Investigation of the potential regulatory role of the 5′-flanking sequence awaits future research efforts.

Also of note, a search of the 3′-noncoding sequence from gcD (AF041818) identified a putative polyadenylation signal motif (data not shown), the authenticity of which has yet to be verified. No significant matches with non-CD39 genes were found.

References

Chadwick BP, Frischauf A-M: Cloning and mapping of a human and mouse gene with homology to ecto-ATPase genes. Mammal Genome 8:668–672 (1997).

Copeland NG, Gilbert DJ, Li K, Sawamura D, Giudice GJ, Chu M-L, Jenkins NA, Uitto J: Chromosomal localization of mouse bullous pemphigoid antigens. BPAG1 and BPAG2: identification of a new region of homology between mouse and human chromosomes. Genomics 15:180–181 (1993).

Copeland NG, Jenkins NA: Development and applications of a molecular genetic linkage map of the mouse genome. Trends Genet 7:113–118 (1991).

Gray IC, Fallowfield J, Ford S, Nobile C, Volpi EV, Spurr NK: An integrated physical and genetic map spanning chromosome band 10q24. Genomics 43:85–88 (1997).

Handa M, Guidotti G: Purification and cloning of a soluble ATP-diphosphohydrolase (apyrase) from potato tubers *(Solanum tuberosum)*. Biochem biophys Res Commun 218:916–923 (1996).

Jackson IJ: A reappraisal of non-consensus mRNA splice sites. Nucl Acids Res 19:3795–3798 (1991).

Kaczmarek E, Koziak K, Sévigny J, Siegel JB, Anrather J, Beaudoin AR, Bach FH, Robson SC: Identification and characterization of CD39/vascular ATP diphosphohydrolase. J biol Chem 271:33116–33122 (1996).

Kegel B, Braun N, Heine P, Maliszewski CR, Zimmermann H: An ecto-ATPase and an ecto-ATP diphosphohydrolase are expressed in rat brain. Neuropharmacology 36:1189–1200 (1997).

Kirley TL: Complementary DNA cloning and sequencing of the chicken muscle ecto-ATPase. Homology with the lymphoid cell activation antigen CD39. J biol Chem 272:1076–1081 (1997).

Maliszewski CR, Delespesse GJT, Schoenborn MA, Armitage RJ, Fanslow WC, Nakajima T, Baker E, Sutherland GR, Poindexter K, Birks C, Alpert A, Friend D, Gimpel SD, Gayle RB III: The CD39 lymphoid cell activation antigen: molecular cloning and structural characterization. J Immunol 153:3574–3583 (1994).

Marcus AJ, Broekman MJ, Drosopoulos JHF, Islam N, Alyonycheva TN, Safier LB, Hajjar KA, Posnett DN, Schoenborn MA, Schooley KA, Gayle RB, Maliszewski CR: The endothelial cell ecto-ADPase responsible for inhibition of platelet function is CD39. J Clin Invest 99:1351–1360 (1997).

Wang TF, Rosenberg PA, Guidotti G: Characterization of brain ecto-apyrase: evidence for only one ecto-apyrase (CD39) gene. Brain Res Mol Brain Res 47:295–302 (1997).